Algorithms for Big Data Delivery over the Internet of Things

Andreas P. Plageras, Student *Member, IEEE, Kostas E. Psannis, Member, IEEE*

***Abstract*—Enormous amounts of data are growing because of the continuous and increasing use of smart devices which connect, collect, exchange and transfer these large amounts of data. With the convergence of new technologies like Internet of Things (IoT), Wireless Sensor Networks (WSNs), and Cloud Computing (CC) many sectors are benefited. Our research has to do with Big Data Delivery over the IoT. Specifically, in our research we are trying to investigate new communication protocols, new security mechanisms, new efficient, faster, safer, and energy efficient solutions and algorithms for healthcare systems (hospital buildings, rooms, etc.). The aim of our research is to create and propose suitable algorithms for efficient transmission of health big data, for management issues, for the analysis of IoT data and for security solutions. Our current position is the experimentation with IoT devices, IoT health data, big data analytics, and healthcare systems. In this paper we present some of the challenges and the issues of Big Data. We also present work that has been done by other researchers. Finally, is presented an open source tool for experimentation, so that we can view the system in detail. Such a tool is the Cooja emulator which is part of the Contiki OS, and is presented also in this paper. As future work we manage to implement such a healthcare system, provide more efficient algorithms, and investigate new protocols for more efficient and secure transmissions of the sensitive health data. By these improvements we are going to provide better healthcare and faster diagnosis to everybody.**

***Index Terms*—Big Data; IoT; WSN; algorithms; security; Cooja; Contiki OS;**

I. Introduction

IN recent years, according to surveys, the emergence of the new technologies and the Internet of Things (IoT) has shown an explosive growth of data. A huge number of networked devices (sensors, actuators, etc.) around the world collect different types of data (environmental, geographical, accounting, etc.). Then, the IoT devices transmit the collected data so that they can be stored, processed, and analyzed. Also, studies have shown that by 2030 approximately one trillion sensors will be connected, which will collect and transfer large amounts of data. Therefore, there is urgent need for adoption of big data (big data) and IoT applications. These technologies are interdependent and should be developed jointly. Specifically, on the one hand, we have the wide spread of IoT, leading to the increase of data, thus providing the opportunity for application and development of big data. On the other hand, the implementation of IoT big data accelerates research advances and business models of IoT [1, 11].

In other words, people in their daily life come in touch with countless devices and other technological advances. These devices are connected with each other and form a network, which leads us to the IoT. This new technological trend is accompanied by countless development and improvement expectations in all areas. Also, this new technological trend comes along with concerns about the security and the violation of privacy.

The general purpose of this work is to help healthcare, so as to become much more accessible and much faster for everyone and everywhere. In this way, with sensors, actuators, cameras, and other components everyone can monitor his/her health, so that the competent doctors and the authorized medical personnel can provide the appropriate care.

The aim of our research is to create and propose, after study and experimentation, suitable algorithms for dealing with the efficient delivery of health big data (delivery), in and out of the network (and to the internet and the cloud), various management issues, the analysis of the IoT-based large-scale data (IoT-Big Data), and security solutions.

The next section (Section II) of this paper is the literature review and the current trends. Specifically, we present some of the technologies that will be involved and converged into an entire communication and automation healthcare system. Moreover, we have the section (Section III) where are presented some challenges and issues of Big Data [13, 15, 16]. Then, we have the methodology section (Section IV) in which we talk about the current research conditions and the future work we are planning to do. Even though, we have the section (Section V), in which are presented some of the results that we could extract from the Cooja emulator. Finally, we conclude the paper in the last section (Section VI).

II. Literature Review & Current Trends

Starting with the "IoT" which is an important and topical issue in the technology industry, and we can say that is the evolution of the Internet, of the computing, of the information, and of the communication systems. We are in a period in which all objects, products, sensors, services and various systems tend to be networked and automated, and offer new features that were not available so far. The culmination of this effort is the IoT which is promising and will literally

Andreas P. Plageras is with the Department of Applied Informatics, University of Macedonia, Greece (e-mail: a.plageras@uom.edu.gr).

K. E. Psannis is with the Department of Applied Informatics, University of Macedonia, Greece (corresponding author, phone: +302310891737; e-mail: kpsannis@uom.edu.gr).

2378-1971/17 $31.00 © 2017 IEEE
DOI 10.1109/CBI.2017.27

transform many aspects of today's lifestyle [5, 6]. One such example is the vision of "Smart Homes" and "Smart Buildings", which will give us greater security and energy efficiency. Due to this vision, we head through networked devices (Internet-enabled), devices capable of managing energy, and home automation devices. Furthermore, cooperation IoT devices with built-in sensors and actuators, devices worn (wearable), health devices with internet connection, and health monitoring devices are transforming the way of delivering healthcare services. Also, these devices offer innovations and applications which will radically change the way of treatment.

Moreover, we deal with the expansion of the internet by computer networks, fixed and mobile communication networks, and the IoT which brings to light the "Wireless Sensor Networks" (WSN). This technology is based on IoT [7]. WSNs are as we understand by the term, networks which consist of sensor nodes and actuators. These devices are connected via a wireless connection (links) and are used to monitor, collect, and measure various environmental and natural data [7].

Progress of health monitoring devices, is the beginning of a new era, where health data collected by sensors will be transmitted and analyzed by healthcare providers. In this new era, the diffusing flow of health data will improve diagnosis and will provide accurate health data, sorter treatment, and will also reduce the difficulties faced by patients before and after their treatment [8].

With the recent developments in WSNs, there are many application areas in the health sector. Using sensors worn, as well as, other types of sensors, patients can be monitored. In healthcare, the monitoring can be done with or without the consent of the patient. From every point, the problems that have to be addressed are the security problems, the privacy, and other related matters [9].

There are too many researchers who study and implement solutions and complete healthcare monitoring systems. Such a healthcare monitoring system for hospitals, which uses WSNs, is the one presented by Media Aminian and Hamid Reza Naji in [10]. In particular, they presented a system for monitoring physiological parameters. These parameters that are detected by the sensors, located in the body of the patient and finally have been contributed a WBAN (Wireless Body Area Network) may be the blood pressure, the heart rate, and so on. Thus, if the system detects changes and conditions which are not normal, sends a notification SMS or e-mail to the responsible doctors for information that can be provided and for the early diagnosis and treatment of the patients.

There are much many schemes like this one which provide different health services and facilitate patient's care. Such a system for monitoring the health of patients at home is planned to be designed in the future by combining all these new technologies that have been analyzed and studied.

III. CHALLENGES AND ISSUES

The rapid growth of data, due to the huge number of networked sensors around the world that collect and transmit, for storage and processing, different types of data, bring enormous challenges for the acquisition, the storage, the management, and the data analysis. Min Chen, Shiwen Mao, and Yunhao Liu discussed in their research [1] some of the key challenges. These concerns:

- the representation of the health data, so as to create more comprehensible health data for computational analysis and interpretation of the user,
- managing the health data life cycle with the development of mechanisms which will be the selection of data that will be stored and those that will be rejected,
- reducing the redundancy which exists in the health data and the compression of these data,
- for the detailed mechanisms,
- the storage capacity and storage media,
- the confidentiality of health big data,
- the management of energy,
- the scalability and consumability, as well as, analytical systems and analytical algorithms should be adapted to different and more complex data sets,
- the collaboration between scientists from different disciplines and development of architectures of large e-health data networks to help these scientists.

Researchers also have some problems, such us fundamental problems of large data (definitions and models), the standardization of large amounts of data, and the development of computational methods of large data.

Continuing with the challenges, we will go to those that have to do with the analysis of large amounts of data (big data analytics). Specifically, researchers at [2], talk about the need that exists for speed and quality of data, and the need for larger storage spaces, such as the 'cloud'. Cloud Computing technology is considered the future technology that will provide storage spaces and mechanisms for data analysis. Also, researchers will need experts on the subject to educate new entrants in this field, because Big Data is a new research field. The integration of different data types is also a challenge.

In addition, in [3] researchers discuss for open issues and challenges of mining large data. One of the important issues is the protection of personal data. To produce information by extracting data requires personal information. The privacy is lost as well, through the social media where an individual's personal data can be extracted. Thus, researchers and scientists should consider in depth and create new tools to address these issues. Also, another important issue mentioned in this research is the user interaction by using feedback/guidance allowed during data mining, so as to enable the user to visualize, interpret and evaluate the results, whether they are finished or intermediate.

Furthermore, we should mention that, since we are talking about Big Data, words such as Hadoop, maps reduce, and Hadoop Distributed File System (HDFS) should be recognized by the general public. [4] More specifically, Hadoop is a free framework in a distributed computing environment which supports the processing of large data, increases the data transfer rate, and also, is fault-tolerant, scalable and flexible. The Hadoop reduction map (Hadoop map reduce) is a frame

which is used for processing and generating large data sets, with a parallel and distributed algorithm in a cluster. The HDFS is a file system that is used by all nodes in a Hadoop cluster (Hadoop cluster) for storing data. More specifically, the connecting HDFS file systems, so as to create a large file system. To have fault tolerance, improves the reliability of the data copied to multiple sources.

IV. METHODOLOGY

Initially, there will be a complete mastery of the Big Data technologies, the Internet of Things, the Cloud Computing, the Wireless Sensor Networks, and general knowledge about the rest of the technologies and the communication protocols that will be involved in this research [12, 14, 17-29]. Such technologies may be the Building Management Systems (BMS) and the Building Automation Systems (BAS), which will allow the management of an entire room, a building, or even a city and provide automation capabilities.

The healthcare sector can be benefited by these technologies in combination with other security mechanisms, protocols, and algorithms [12, 14, 17-29]. Such a security mechanism is provided in [30] and is called Named Data Networking (NDN). With this mechanism many risks have been overtaken and the access to the data is inspected. Also, as privacy solutions will be used the Advanced Encryption Standard (AES), Datagram Transport Layer Security (DTLS) mechanisms, or other security trends and novel investigations.

The protocols that will certainly, participate are the Hypertext Transfer Protocol (HTTP), the Constrained Application Protocol (CoAP), the User Datagram Protocol (UDP), the Transmission Control Protocol/Internet Protocol (TCP/IP), the IPv6 over Low-power Personal Area Network (6LoWPAN), the IEEE 802.15.4 or newer, the Message Queuing Telemetry Transport (MQTT) protocol, and other similar or new investigated [11]. In the following Table I a presentation of the layers and some of the protocols that can be used in each layer are presented.

TABLE I.
The layers and the protocols can be used in each layer.

Layers	Sub-layers	Protocols
Physical	-	IEEE 802.15.4
Data Link	MAC	IEEE 802.15.4
	6LoWPAN	6LoWPAN
Network	-	IPv6/HTTP
Transport	-	CoAP/UDP
Application	-	CoAP

The architecture of the system we plan to implement will be an entire automated system of sensor and body area networks that will be built in the appropriate topologies for example the Hybrid Network Topology of the following Figure 1 [11].

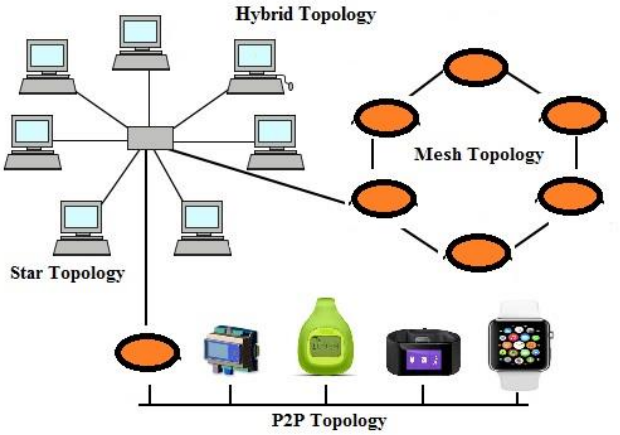

Figure 1. Hybrid Network Topology.

After a study of the challenges, the issues, and other similar researches, we are going to make a convergence of technologies, protocols, mechanisms, and algorithms, to provide better solutions. To make such a thing possible we will need simulation tools like the Cooja emulator which is part of the Contiki OS and will be useful for experimentation. The Contiki is designed to support IP networking in low-power network connections. It is implemented in C language and uses a structured environment that can be applied to the most of the platforms. In the Cooja emulator we can simulate wireless sensor networks. It is a flexible java simulator that supports C for application software development. The parts of the simulation environment can be changed by the application developers without changing the Cooja code. This means that new partitions, such as plugins, interfaces or remodeling existing segments can be added to the system. With these advantages of Cooja, we can implement many simulations with different system conditions and settings, such as different packet data rates, different MAC protocols, and different topology of the network.

With such tools we can simulate in real time our sensor networks. As a result of the experiments will be the more detailed view of our network and the conditions in it. In "Figure 2" below, the environment of the Cooja emulator and the network simulated can be seen.

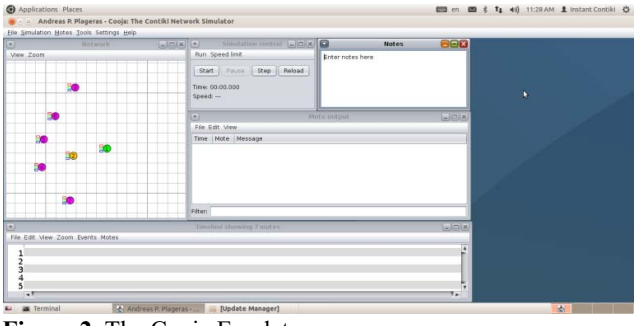

Figure 2. The Cooja Emulator.

V. Using the Cooja Emulator

Cooja has sophisticated tools for collecting data from nodes to help the user draw conclusions about the simulation of the wireless network. Once the simulation has been completed, a large number of data will be available for analysis. Most of the information is graphical and has to do with topology, with sensors and in particular with information about temperature, humidity, battery, etc. We also receive information about our network metrics, such as for the packets received, the packets lost during the simulation every five minutes or per node, metrics for the router, etc. Furthermore, there are also provided information about the energy consumed at specific time periods, the average values, the history of the simulation, etc.

In the following "Figure 3" we can see the communications between the sensor devices (nodes) in a network. We can also see the results after the execution of ping to the IPv6 address of the sensor selected, in this case the node selected is the router (node 1). Moreover, in "Figure 4" are shown the packets that have been transmitted, received, and lost by the selected node.

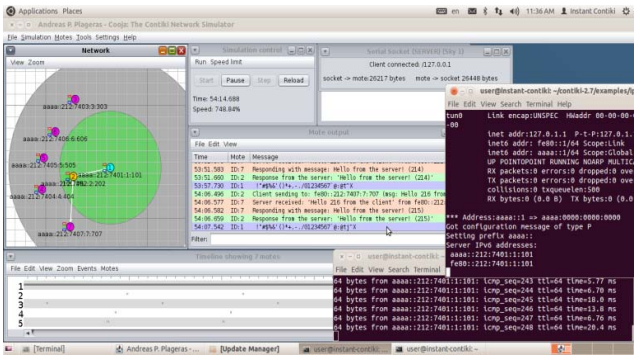

Figure 3. Transmission in our network.

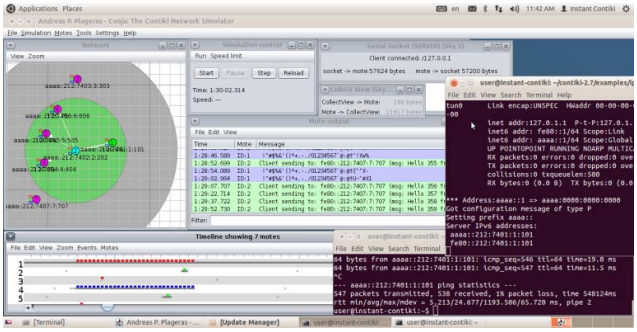

Figure 4. The packets received and lost (over time).

For cloud analytics and further experimentation we are possibly going to use open source cloud simulators such as the CloudSim Java-based simulation toolkit, the CloudAnalyst, the GreenCloud, the iCanCloud, and so on.

Once the study is completed, we will pass the experimental level, in which there will be proposed algorithms for the delivery of big data, for security and so on. Finally, after the testing part is completed, we will check, verify, and present the results.

VI. Conclusion

With the evolution of the technologies of Big Data, IoT, and Cloud Computing, significant advances have been made in the healthcare sector, and not only [12, 14, 17-29]. In this paper, a wide range of research has been studied in the field of IoT, CC, and other typical technologies of effective healthcare systems. Also, in this paper we have presented some of the challenges and the issues of these large-amounts of IoT data. Moreover, we have presented some of the technologies, the communication protocols, the security mechanisms, and other techniques that will be involved in our research. Even though, we continue to study about simulation tools, that will be useful (e.g. Contiki OS or CloudSim) for our experiments. The limitations of this research are the limited energy and the privacy. In particular, the problem of limiting the devices to energy needs must be improved so that there can be used more sophisticated and complex security mechanisms. In the future we want to investigate new communication protocols, algorithmic solutions for the delivery of big amounts of e-health data, security solutions, and efficient systems. This research will be a start point for designing better and more efficient healthcare systems.

Appendix

Recommendation Letter

Dear Selection Committee,

I am writing this letter to strongly recommend Mr. Plageras for your conference. I know Mr. Plageras because we had cooperated in his dissertation for his Master's degree in Technology Management, University of Macedonia. I met with him and outlined a project. I gave him some background reading at our first meeting. By the time of our second meeting he had read what I had given him and prepared a two-page project description.

During his time in School of Technology Management, University of Macedonia, Mr. Plageras demonstrated a good work ethic and interpersonal skills. We outlined a scope of work to be completed, and he successfully completed that work in the time required. He put in extra hours as necessary in order to meet specific deadlines that I set. I teamed him up with another student to work on the project. He seemed to work well with the other student, and I found him very personable. Mr. Plageras put in sufficient work to be a co-author on a manuscript.

In summary, Mr. Plageras is clearly the best student I have worked with in the last 10 years. I would very much like him match to our residency program. Even though, I hope he stays here, I think he would be an outstanding asset to your conference. I give him my highest recommendation.

Sincerely,
Dr. Konstantinos E. Psannis

REFERENCES

[1] Min Chen, Shiwen Mao, and Yunhao Liu, "Big Data: A Survey", Mobile Network Applications, Vol. 19, Pages: 171-209, 2014.

[2] Sunita Sahu and Yugchhaya Dhote, "A Study on Big Data: Issues, Challenges and Applications", International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), 2016.

[3] Abhinav Kathuria, "Issues and Challenges in the Era of Big Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 6, Issue 6, June 2016.

[4] Reena Singh and Kunver Arif Ali, "Challenges and Security Issues in Big Data Analysis", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 1, January, 2016.

[5] Internet Society, "The Internet of Things: An Overview". INTERNETSOCIETY.ORG, October 2015.

[6] K. E. Psannis, S. Xinogalos, and A. Sifaleras, "Convergence of Internet of Things and Mobile Cloud Computing", Systems Science & Control Engineering (An open access journal), Vol. 14, No. 1, Pages: 476-483, Department of Applied Informatics, School of Information Sciences, University of Macedonia, Greece, 2014. http://www.tandfonline.com/doi/full/10.1080/21642583.2014.913213

[7] Fabian Nack, Institute of Computer Science (ICS), "An Overview on Wireless Sensor Networks", 2009.

[8] Yuan Zhang, Limin Sun, Houbing Song, και Xiaojun Cao, "Ubiquitous WSN for Healthcare: Recent Advances and Future Prospects", IEEE Internet of Things Journal, Vol. 1, No. 4, August 2014.

[9] Moshaddique Al Ameen και Kyung-sup Kwak, "Social Issues in Wireless Sensor Networks with Healthcare Perspective", The International Arab Journal of Information Technology, Vol. 8, No. 1, January 2011.

[10] Media Aminian και Hamid Reza Naji, "A Hospital Healthcare Monitoring System Using Wireless Sensor Networks", Journal on Health and Medical Informatics, Vol. 4, Issue: 2, 2013.

[11] **Andreas P. Plageras**, Kostas E. Psannis, Yutaka Ishibashi, Byung-Gyu Kim, "IoT-based Surveillance System for Ubiquitous Healthcare", 42nd Annual Conference of the IEEE Industrial Electronics Society, 24/10/2016 - 27/10/2016.Jay Lee, Behrad Bagheri, and Hung-An Kao, "Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics", International Conference on Industrial Informatics (INDIN), July 2014.

[12] Yixian Zheng, Wenchao Wu, Yuanzhe Chen, Huamin Qu, and Lionel M. Ni, "Visual Analytics in Urban Computing: An Overview", IEEE Transactions on Big Data, Vol. 2, No. 3, July-September 2016.

[13] Xiaolong Jin, Benjamin W. Wah, Xueqi Cheng, and Yuanzhuo Wang, "Significance and Challenges of Big Data Research", Elsevier, Big Data Research, Vol. 2, pages: 59-64, 2015.

[14] Panagiotis D. Diamantoulakis, Vasileios M. Kapina, and George K. Karagiannidis, "Big Data Analytics for Dynamic Energy Management in Smart Grids", Elsevier, Big Data Research, Vol. 2, Pages: 94-101, 2015.

[15] Michele De Gennaro, Elena Paffumi, and Giorgio Martini, "Big Data for Supporting Low-Carbon Road Transport Policies in Europe: Applications, Challenges and Opportunities", Elsevier, Big Data Research, Vol. 6, Pages: 11-25, 2016.

[16] Pekka Paakkonen and Daniel Pakkala, "Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems", Elsevier, Big Data Research, Vol. 2, Pages: 166-186, 2015.

[17] Priyank Jain, Manasi Gyanchandani, and Nilay Khare, "Big data privacy: a technological perspective and review", Springer, Journal of Big Data, Vol. 3, No. 25, 2016.

[18] Yi Wang, Yi Li, Momiao Xiong, Yin Yao Shugart, and Li Jin, "Random bits regression: a strong general predictor for big data", Springer, Big Data Analytics, Vol. 1, No. 12, 2016.

[19] Uthayanath Suthakar, Luca Magnoni, David Ryan Smith, Akram Khan, and Julia Andreeva, "An efficient strategy for the collection and storage of large volumes of data for computation", Springer, Journal of Big Data, Vol. 3, No. 21, 2016.

[20] Salvador Garcia, Sergio Ramirez-Gallego, Julian Luengo, Jose Manuel Benitez, and Francisco Herrera, "Big data preprocessing: methods and prospects", Springer, Big Data Analytics, Vol. 1, No. 9, 2016.

[21] Monteith Scott and Glenn Tasha, "Automated Decision-Making and Big Data: Concerns for People With Mental Illness", Current Psychiatry Reports, Vol. 18, Issue: 112, December 2016.

[22] El Majdouli Mohamed Amine, Rbouh Ismail, Bougrine Saad, El Benani Bouazza, and El Imrani Abdelhakim Ameur, "Fireworks algorithm framework for Big Data optimization", Memetic Computing, Vol. 8, Issue: 4, Pages: 333-347, December 2016.

[23] Farid Dewan Md, Al-Mamun Mohammad Abdullah, Manderick Bernard, and Nowe Ann, "An adaptive rule-based classifier for mining big biological data", Expert Systems with Applications, Vol. 64, Pages: 305-316, December 2016.

[24] Ghosh Goldina, Banerjee Soumya, and Yen Neil Y., "State transition in communication under social network: An analysis using fuzzy logic and Density Based Clustering towards big data paradigm", Future Generation Computer Systems, The International Journal of EScience, Vol. 65, Pages: 207-220, December 2016.

[25] Onose Alexandru, Carrillo Rafael E., Repetti Audrey, McEwen Jason D., Thiran Jean-Philippe, Pesquet Jean-Christophe, and Wiaux Yves, "Scalable splitting algorithms for big-data interferometric imaging in the SKA era", Monthly Notices of the Royal Astronomical Society, Vol. 462, Issue: 4, Pages: 4314-4335, November 2016.

[26] Zhang De-Gan, Zhou Shan, Chen Jie, and Liu Si, "New mixed adaptive detection algorithm for moving target with big data", Journal of Vibroengineering, Vol. 18, Issue: 7, Pages: 4705-4719, November 2016.

[27] Ow Ghim Siong and Kuznetsov Vladimir A., "Big genomics and clinical data analytics strategies for precision cancer prognosis", Scientific Reports, Vol. 6, November 2016.

[28] Xiong Wen, Yu Zhibin, Eeckhout Lieven, Bei Zhengdong, Zhang Fan, and Xu Chengzhong, "ShenZhen transportation system (SZTS): a novel big data benchmark suite", Journal of Supercomputing, Vol. 72, Issue: 11, Pages: 4337-4364, November 2016.

[29] Rodriguez-Jimenez Jose, Cordero Pablo, Enciso Manuel, and Mora Angel, "Data mining algorithms to compute mixed concepts with negative attributes: an application to breast cancer data", Mathematical Methods in the Applied Sciences, Vol. 39, Issue: 16, Pages: 4829-4845, November 2016.

[30] Wentao Shang, Qiuhan Ding, Alessandro Marianantoni, Jeff Burke, and Lixia Zhang, "Securing Building Management Systems Using Named Data Networking", IEEE Network, May/June 2014.